

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-250914

(43)Date of publication of application : 14.09.2000

(51)Int.Cl.

G06F 17/28

G06F 17/27

(21)Application number : 11-053139

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

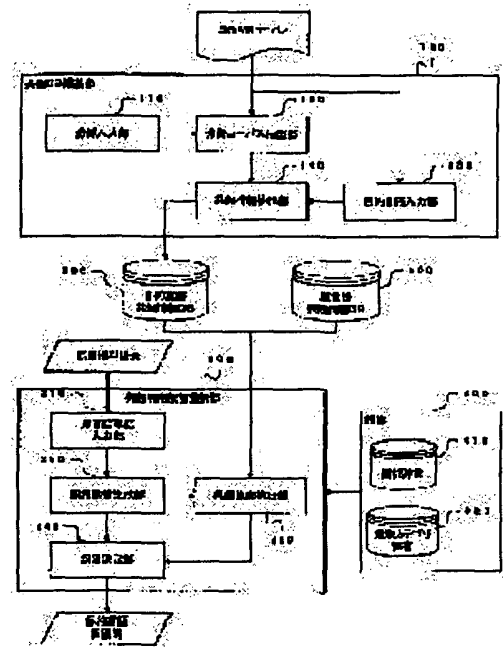
(22)Date of filing : 01.03.1999

(72)Inventor : ASANOMA NAOIKI
NAKAIWA HIROMI(54) MACHINE TRANSLATION METHOD AND DEVICE AND RECORDING MEDIUM
RECORDING MACHINE TRANSLATION PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a machine translation method/device which can suppress the scale of a co-occurrence information data base and also to obtain a most likelihood translation equivalent string against an original language word string without using an analyzed corpus.

SOLUTION: The word pairs occurring concurrently in a decided range of a sentence of an object language corpus and the co-occurrence frequency information on these word pairs are stored in an object language co-occurrence information data base 300. A translation equivalent candidate generation part 220 inputs an original language word string to be translated and retrieves the translation equivalent candidates of an object language of every original language word included in the original language word string by means of a translation dictionary 410. A co-occurrence intensity detection part 230 retrieves the relative words and near-synonyms of every object language translation equivalent by means of the dictionary 410 or a semantic category dictionary 420 and calculates the co-occurrence intensity of a translation equivalent candidate pair against the original language word by means of the data base 300. A translation equivalent decision part 240 selects an object language translation equivalent against the original language word according to the co-occurrence intensity.



LEGAL STATUS

[Date of request for examination] 21.12.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-250914
(P2000-250914A)

(43) 公開日 平成12年9月14日 (2000.9.14)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード(参考)
G 0 6 F 17/28		G 0 6 F 15/38	C 5 B 0 9 1
17/27			E

審査請求 未請求 請求項の数43 O L (全 16 頁)

(21) 出願番号 特願平11-53139

(22) 出願日 平成11年3月1日 (1999.3.1)

(71) 出願人 000004226

日本電信電話株式会社
東京都千代田区大手町二丁目3番1号

(72) 発明者 麻野間 直樹

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 中岩 浩巳

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 100088328

弁理士 金田 暢之

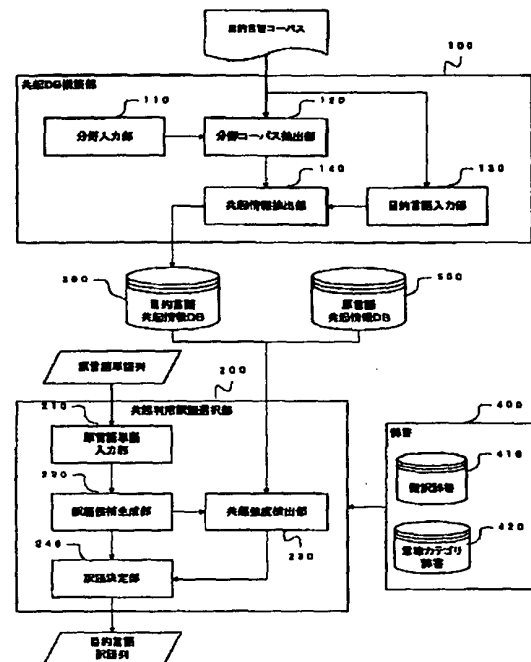
Fターム(参考) 5B091 AA15 CA02 CA22 CC05 CC18
EA17

(54) 【発明の名称】 機械翻訳方法、装置、および機械翻訳プログラムを記録した記録媒体

(57) 【要約】

【課題】 機械翻訳において、単語間依存情報が付加されたコーパスを用いることなく、訳語選択処理の不自然な訳語列出力と、コーパスにおける共起現象の希薄さの問題点を解決する。

【解決手段】 目的言語コーパスの文において、定めた範囲に同時に出現する単語対とその共起頻度情報が目的言語共起情報データベース300に蓄積される。訳語候補生成部220は、翻訳対象の原言語単語列を入力し、翻訳辞書410を用いてその原言語単語列中の各原言語単語の目的言語の訳語候補を検索する。共起強度検出部230は、翻訳辞書410または意味カテゴリ辞書420を用いて各目的言語訳語の関連語、類義語を検索し、共起情報データベース300を用いて原言語単語に対する訳語候補対の共起強度を計算する。訳語決定部240は共起強度を用いて、原言語単語に対する目的言語訳語列を選択する。



1

【特許請求の範囲】

【請求項1】 目的言語の文の集合からなる目的言語コーパス中の文において、定めた範囲内に同時に出現する単語対とその共起頻度情報を共起情報データベースに蓄積する目的言語共起情報データベース構築方法。

【請求項2】 目的言語の文の集合からなる目的言語コーパスを入力する目的言語入力ステップと、該目的言語コーパス中の文において、定めた範囲内に同時に出現する単語対とその共起頻度情報を前記共起情報データベースに蓄積する共起情報抽出ステップを含む、請求項1記載の方法。

【請求項3】 前記目的言語共起情報データベースは、目的言語の、品詞情報が付与された単語対である品詞タグ付き単語対とその共起頻度情報からなるエントリの集合を保持する、請求項1記載の方法。

【請求項4】 目的言語の各単語にその単語の品詞情報が付与された文の集合からなる目的言語コーパスを入力する目的言語入力ステップと、該目的言語コーパス中の文において、定めた範囲内に同時に出現する品詞タグ付き単語対とその共起頻度情報を前記共起情報データベースに蓄積する共起情報抽出ステップを含む、請求項3記載の方法。

【請求項5】 前記共起情報抽出ステップは、品詞タグ付き単語対を収集する際に、あるキーとなる単語であるキー単語に対して共起する単語である共起単語を品詞別に集計し、各共起単語について品詞別に何番目に近いかを示す品詞別共起順位を抽出し、該キー単語と該共起単語の対、および該品詞別共起順位別に共起頻度情報を前記共起情報データベースに蓄積する、請求項4記載の方法。

【請求項6】 分野情報を入力する分野入力ステップと、前記目的言語コーパスから該分野情報に関連する文の集合を抽出する分野コーパス抽出ステップをさらに含み、前記共起情報抽出ステップにおいて、該分野コーパス入力ステップで抽出された文中で目的言語の共起頻度情報を収集する、請求項2、4、5のいずれか1項記載の方法。

【請求項7】 原言語単語と目的言語訳語候補の対訳関係の集合を保持する翻訳辞書を用いて、翻訳対象の原言語列中の各原言語単語の目的言語訳語候補を検索する訳語候補生成ステップと、該目的言語訳語候補と請求項1から6のいずれか1項に記載の共起情報データベースのエントリとを照合して、原言語単語に対する該目的言語訳語候補対の共起強度を計算する共起強度検出ステップと、前記共起強度を用いて、前記原言語単語に対する目的言語訳語列を選択する訳語決定ステップを含む機械翻訳方法。

【請求項8】 前記原言語単語列を入力するステップを

2

さらに含む請求項7記載の方法。

【請求項9】 前記共起強度検出ステップにおいて、品詞情報を含んだ請求項3記載の共起情報データベースを利用して、該品詞情報を利用して原言語単語に対する目的言語訳語候補対の共起強度を計算する、請求項7または8記載の方法。

【請求項10】 前記共起強度検出ステップにおいて、請求項5記載の該品詞別共起順位と該共起頻度情報を加味して共起強度を計算する請求項9記載の方法。

【請求項11】 前記翻訳辞書は、目的言語単語と該目的言語単語の変化形、派生語、同義語である目的言語単語関連語の対応関係の集合を保持し、前記共起強度検出ステップにおいて、該翻訳辞書を参照して、各目的言語訳語候補と前記共起情報データベースのエントリ中の目的言語単語対を照合する際に、目的言語単語どうしだけでなく該目的言語単語の目的言語単語関連語とも照合して共起強度を計算する、請求項7ないし10のいずれか1項記載の方法。

【請求項12】 前記共起強度検出ステップにおいて、前記翻訳辞書を参照して、各目的言語訳語候補と前記共起情報データベースのエントリ中の目的言語単語対を照合する際に、目的言語単語どうしだけでなく、目的言語の単語とその単語の意味を代表する意味カテゴリとの対応関係の集合から構成される意味カテゴリ辞書を参照して、該目的言語単語の意味カテゴリどうしとも照合して共起強度を計算するステップを含む、請求項7ないし11のいずれか1項記載の方法。

【請求項13】 前記翻訳辞書は、前記原言語単語に対して、選択すべき目的言語訳語候補を決定するための条件を記述したルール制約情報を含み、前記訳語候補生成ステップにおいて、該翻訳辞書を参照して、目的言語訳語候補にルール制約情報を付与し、前記訳語決定ステップにおいて、目的言語訳語候補に付与された該ルール制約情報に矛盾しない目的言語訳語候補を選択する請求項7ないし12のいずれか1項記載の方法。

【請求項14】 前記訳語候補生成ステップにおいて、該訳語候補生成ステップで用いる前記翻訳辞書に加えて、原言語単語と目的言語訳語候補の対訳関係の集合を保持する別の翻訳辞書も参照して前記原言語単語列中の各原言語単語の目的言語訳語候補を検索する、請求項7ないし13のいずれか1項記載の方法。

【請求項15】 前記共起強度検出ステップにおいて、原言語の単語対とその共起頻度情報からなるエントリの集合を保持する原言語共起情報データベースを用いて、原言語単語対に対する原言語における共起強度をもとに、対応する訳語候補共起強度に重み付けを行う、請求項7ないし14のいずれか1項記載の方法。

【請求項16】 目的言語の文の集合からなる目的言語コーパスを入力する目的言語入力手段と、

該目的言語コーパス中の文において、定めた範囲内に同時に出現する単語対とその共起頻度情報を共起情報データベースに蓄積する共起情報抽出手段を有する目的言語共起情報データベース構築装置。

【請求項17】 前記目的言語共起情報データベースは、目的言語の品詞タグ付き単語対とその共起頻度情報からなるエントリの集合を保持する、請求項16記載の装置。

【請求項18】 前記目的言語入力手段は、目的言語の、各単語にその単語の品詞情報が付与された文の集合からなる目的言語コーパスを入力し、前記共起情報抽出手段は、目的言語コーパス中の文において、定めた範囲内に同時に出現する品詞タグ付き単語対とその共起頻度情報を前記共起情報データベースに蓄積する、請求項17記載の装置。

【請求項19】 前記共起情報抽出手段は、品詞タグ付き単語対を収集する際に、あるキーとなる単語であるキー単語に対して共起する単語である共起単語を品詞別に集計し、各共起単語について品詞別に何番目に近いかを示す品詞別共起順位を抽出し、該キー単語と該共起単語の対、および該品詞別共起順位に共起頻度情報を前記共起情報データベースに蓄積する手段を有する、請求項18記載の装置。

【請求項20】 分野情報を入力する分野入力手段と、前記目的言語コーパスから該分野情報に関連する文の集合を抽出する分野コーパス抽出手段とをさらに有し、前記共起情報抽出手段は、該分野コーパス入力手段で抽出された文中で目的言語の共起頻度情報を収集することを含む請求項16から19のいずれか1項記載の装置。

【請求項21】 原言語単語と目的言語訳語候補の対訳関係の集合を保持する翻訳辞書と、該翻訳辞書を用いて翻訳対象の原言語単語列中の各原言語単語の目的言語訳語候補を検索する訳語候補生成手段と、該目的言語訳語候補と請求項16～20に記載の共起情報データベースのエントリとを照合して、原言語単語に対する該目的言語訳語候補対の共起強度を計算する共起強度検出手段と、前記共起強度を用いて、原言語単語に対する目的言語訳語列を選択する訳語決定手段を有する機械翻訳装置。

【請求項22】 前記原言語単語列を入力する原言語単語列入力手段をさらに有する、請求項21記載の装置。

【請求項23】 前記共起強度検出手段は、該品詞情報を含んだ前記共起情報データベースを利用して、該品詞情報を利用して原言語単語に対する目的言語訳語候補対の共起強度を計算する、請求項21または22記載の装置。

【請求項24】 前記共起強度検出手段は、該品詞別共起順位と該共起頻度情報を加味して共起強度を計算する手段を含む請求項21から23のいずれか1項記載の装

置。

【請求項25】 前記翻訳辞書は、目的言語単語と該目的言語単語の変化形、派生語、同義語である目的言語単語関連語の対応関係の集合を保持し、前記共起強度検出手段は、該翻訳辞書を参照して、各目的言語訳語候補と前記共起情報データベースのエントリ中の目的言語単語対を照合する際に、目的言語単語どうしだけでなく該目的言語単語の目的言語単語関連語とも照合して共起強度を計算する手段を含む、請求項21ないし24のいずれか1記載の装置。

【請求項26】 目的言語の単語とその単語の意味を代表する意味カテゴリとの対応関係の集合から構成される意味カテゴリ辞書を有し、前記共起強度検出手段は、前記翻訳辞書を参照して、各目的言語訳語候補と前記共起情報データベースのエントリ中の目的言語単語対を照合する際に、目的言語単語どうしだけでなく、該意味カテゴリ辞書を参照して、該目的言語単語の意味カテゴリどうしとも照合して共起強度を計算する手段を含む、請求項21ないし25のいずれか1項記載の装置。

【請求項27】 前記翻訳辞書は、前記原言語単語に対して、選択すべき目的言語訳語候補を決定するための条件を記述したルール制約情報を含み、前記訳語候補生成手段は、該翻訳辞書を参照して、目的言語訳語候補にルール制約情報を付与する手段を含み、前記訳語決定手段は、目的言語訳語候補に付与された該ルール制約情報に矛盾しない目的言語訳語候補を選択する手段を含む、請求項21ないし26のいずれか1項記載の装置。

【請求項28】 前記訳語候補生成手段は、該訳語候補生成手段で用いる前記翻訳辞書に加えて、原言語単語と目的言語訳語候補の対訳関係の集合を保持する別の翻訳辞書も参照して前記原言語単語列中の各原言語単語の目的言語訳語候補を検索する手段を含む、請求項21ないし27のいずれか1項記載の装置。

【請求項29】 原言語の単語対とその共起頻度情報からなるエントリの集合を保持する原言語共起情報データベースをさらに有し、前記共起強度検出手段は、該原言語共起情報データベースを用いて、原言語単語対に対する原言語における共起強度をもとに、対応する訳語候補共起強度に重み付けを行う手段を含む、請求項21ないし28のいずれか1項記載の装置。

【請求項30】 目的言語の文の集合からなる目的言語コーパスを入力する目的言語入力手段と、該目的言語コーパス中の文において、定めた範囲内に同時に出現する単語対とその共起頻度情報を共起情報データベースに蓄積する共起情報抽出手段をコンピュータに実行させるための共起情報抽出プログラムを記録した記録媒体。

【請求項31】 前記目的言語共起情報データベース

5

は、目的言語の、品詞情報が付与された単語対である品詞タグ付き単語対とその共起頻度情報からなるエントリの集合を保持する、請求項30記載の記録媒体。

【請求項32】 目的言語の、各単語毎にその単語の品詞情報が付与された文の集合からなる目的言語コーパスを入力する目的言語入力手順と、該目的言語コーパス中の文において、定めた範囲内に同時に出現する品詞タグ付き単語対とその共起頻度情報を共起情報データベースに蓄積する共起情報抽出手順をコンピュータに実行させるための共起情報抽出プログラムを記録した記録媒体。

【請求項33】 前記共起情報抽出手順は、品詞タグ付き単語対を収集する際に、あるキーとなる単語であるキー単語に対して共起する単語である共起単語を品詞別に集計し、各共起単語について品詞別に何番目に近いかを示す品詞別共起順位を抽出し、該キー単語と該共起単語の対、および該品詞別共起順位別に共起頻度情報を前記共起情報データベースに蓄積する、請求項32に記載の記録媒体。

【請求項34】 分野情報を入力する分野入力手順と、前記目的言語コーパスから該分野情報に関連する文の集合を抽出する分野コーパス抽出手順をさらに含み、前記共起情報抽出手順において、該分野コーパス入力手順で抽出された文中でも目的言語の共起頻度情報を収集する、請求項30、32、33のいずれか1項記載の記録媒体。

【請求項35】 原言語単語と目的言語訳語候補の対訳関係の集合を保持する翻訳辞書を用いて翻訳対象の原言語単語列中の各原言語単語の目的言語訳語候補を検索する訳語候補生成手順と、
該目的言語訳語候補と請求項30から34のいずれか1項に記載の共起情報データベースのエントリとを照合して、原言語単語に対する該目的言語訳語候補対の共起強度を計算する共起強度検出手順と、
前記共起強度を用いて、原言語単語に対する目的言語訳語列を選択する訳語決定手順をコンピュータに実行させるための機械翻訳プログラムを記録した記録媒体。

【請求項36】 前記原言語単語列を入力するステップをさらに含む請求項35記載の記録媒体。

【請求項37】 前記共起強度検出手順において、品詞情報を含んだ、請求項31に記載の共起情報データベースを利用して、該品詞情報を利用して原言語単語に対する目的言語訳語候補対の共起強度を計算する、請求項35または36記載の記録媒体。

【請求項38】 前記共起強度検出手順において、請求項31記載の該品詞別共起順位と該共起頻度情報を加味して共起強度を計算する請求項37記載の記録媒体。

【請求項39】 前記翻訳辞書は、目的言語単語と該目的言語単語の変化形、派生語、同義語である目的言語単語関連語の対応関係の集合を保持し、
前記共起強度検出手順において、該翻訳辞書を参照し

6

て、各目的言語訳語候補と前記共起情報データベースのエントリ中の目的言語単語対を照合する際に、目的言語単語どうしだけでなく該目的言語単語の目的言語単語関連語とも照合して共起強度を計算する、請求項35ないし38のいずれか1項記載の記録媒体。

【請求項40】 前記共起強度検出手順において、前記翻訳辞書を参照して、各目的言語訳語候補と前記共起情報データベースのエントリ中の目的言語単語対を照合する際に、目的言語単語どうしだけでなく、目的言語の単語とその単語の意味を代表する意味カテゴリとの対応関係の集合から構成される意味カテゴリ辞書を参照して、該目的言語単語の意味カテゴリどうしとも照合して共起強度を計算する手順を含む、請求項35ないし39のいずれか1項記載の記録媒体。

【請求項41】 前記翻訳辞書は、前記原言語単語に対して、選択すべき目的言語訳語候補を決定するための条件を記述したルール制約情報を含み、
前記訳語候補生成手順において、該翻訳辞書を参照して、目的言語訳語候補にルール制約情報を付与し、
前記訳語決定手順において、目的言語訳語候補に付与された該ルール制約情報に矛盾しない目的言語訳語候補を選択する請求項35ないし40のいずれか1項記載の記録媒体。

【請求項42】 前記訳語候補生成手順において、該訳語候補生成手順で用いる前記翻訳辞書に加えて、原言語単語と目的言語訳語候補の対訳関係の集合を保持する別の翻訳辞書も参照して前記原言語単語列中の各原言語単語の目的言語訳語候補を検索する請求項35ないし41のいずれか1項記載の記録媒体。

【請求項43】 前記共起強度検出手順において、原言語の単語対とその共起頻度情報からなるエントリの集合を保持する原言語共起情報データベースを用いて、原言語単語対に対する原言語における共起強度をもとに、対応する訳語候補共起強度に重み付けを行う請求項35ないし42のいずれか1項記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、目的言語の単語共起情報を利用した訳語選択を行う機械翻訳方法および装置に関する。

【0002】

【従来の技術】辞書やルールを用いたルールベースの機械翻訳システムにおいては、対訳辞書や変換ルールにより、入力文を解析後、翻訳文を生成する処理が行われる。この翻訳文を生成する段階において適切な訳語を選択させる手法としては、例えば、原言語単語の意味カテゴリを判定し、意味カテゴリレベルの変換ルールを用いて訳し分けする方法（白井、横尾、内野、松尾、「日英変換技術と意味辞書」、NTT R&D, Vol. 46, pp.1405-1410, 1997）や、原言語単語の類義語を検索し、変換ル

7

ルの適用可能性を増やして訳し分けする方法（例えば、「機械翻訳装置」特開平5-158970）が知られている。

【0003】さらに、より適切な訳語を選択するための改良手法としては、コーパスから獲得した統計的知識を利用する方法がある。統計的知識による訳語選択を行う手法には、対訳コーパスを利用する方法や、目的言語側の単語共起情報を用いた訳語選択方法が提案されている。

【0004】対訳コーパスを利用する方法としては、例えば、単語対応のとれた対訳コーパスから得られる、翻訳対象単語の意味と原言語の共起単語との統計的知識を用いる語義多義性解消（訳語選択）方法が挙げられる

(P. Brown, S. Della Pietra, V. Della Pietra, R. Mercer, "Word-sense disambiguation using statistical methods", Proceedings of Annual Meeting of the Association for Computational Linguistics, pp. 264-270, 1991)。

【0005】目的言語側の単語共起情報を用いた訳語選択方法の例としては、目的言語コーパスにおいて依存関係のある共起単語の頻度を収集し、これらを用いて訳語二つの組に対する依存関係の強度を選択基準値として、多訳動詞の訳し分けを行う方法が提案されている（野見山、「目的言語の知識を用いた訳語選択とその学習性」、情報処理学会研究会資料、NL86-8、1991）。

【0006】また、目的言語コーパスにおいて共起単語の頻度情報を収集した後、選択基準値は、訳語の組み合わせの一番高い共起確率と二番目に高い共起確率の比の値を基本とし、選択基準値が統計的に有意なしきい値を超えるような訳語組を選択する手法が提案されている。

(Ido Dagan, Alon Itai, "Word sense disambiguation using a second language monolingual corpus", Computational Linguistics, Vol.20, No. 4, pp. 563-596, 1994)。

【0007】

【発明が解決しようとする課題】従来のルールベースの機械翻訳システムにおける訳語選択手法においては、次のような問題点を有している。

【0008】ある原言語単語の訳語を選択する場合、ある原言語単語は、ルールと辞書を用いて目的言語訳語候補を取得し、辞書に記述された訳語選択の優先順位と変換ルールによる制約によって、ある一つの訳語が選択される。その後、入力複数の単語に対して、各々選ばれた訳語を並べて合成（要素合成）する。このようにして得られる翻訳結果は、訳語に生成する際の適切性を十分に考慮していないため、目的言語として不自然な訳語が並びやすいという問題がある。

【0009】また、適切な訳語が機械翻訳システムの翻訳辞書に登録されておらず、目的言語訳語候補の選択肢が十分でない場合がある。

8

【0010】さらに、前記コーパスに基づく訳語選択手法では、単語対応や文対応のとれた対訳コーパスや、目的言語コーパスを用いる場合においても、人手によって正確に付加した単語間依存情報を持つコーパスは、依存情報の付加されていないコーパスよりも人手が困難である。そのため、統計的知識の網羅的な取得が難しいという問題がある。また、目的言語のコーパスを構文解析する場合は、解析失敗による誤った依存情報が、そこから得られる統計的知識に誤った情報として影響してしまうという問題がある。それゆえ、依存情報のない目的言語コーパスの利用を考慮に入れることが望ましい。その上、コーパスを用いる自然言語処理の性質として、コーパスに出現する単語表記に対する共起情報を収集すると、単語の共起現象が希薄になり、収集される個々の共起頻度情報は小さな値をとることが多くなる。結果として、個々の共起頻度情報の信頼性が下がり、このようにして構築した共起情報データベースを用いて訳語選択を行う場合、訳語候補に誤りが混ざりやすいという問題がある。また、原言語におけるルールによる目的言語訳語候補の制約が考慮されていないという問題点や、原言語単語についての共起強度を考慮していないという問題点がある。

【0011】本発明の目的は、単語間依存情報が付加されたコーパスを用いることなく、訳語選択処理の不自然な訳語列出力と、コーパスにおける共起現象の希薄さの問題点を解決した機械翻訳方法、装置、および機械翻訳プログラムを記録した記録媒体を提供することにある。

【0012】

【課題を解決するための手段】本発明の機械翻訳方法は、原言語単語と目的言語訳語候補の対訳関係の集合を保持する翻訳辞書を用いて、翻訳対象の原言語単語列中の各原言語単語の目的言語訳語候補を検索する訳語候補生成ステップと、該目的言語訳語候補と、共起情報データベースのエントリを照合して、原言語単語に対する該目的言語訳語候補対の共起強度を計算する共起強度検出ステップと、前記共起強度を用いて、原言語単語に対する目的言語訳語列を選択する訳語決定ステップを含む。

【0013】本発明の共起情報データベース構築方法は、目的言語の文の集合からなる目的言語コーパス中の文において、定めた範囲内に同時に出現する単語対とその共起頻度情報を共起情報データベースに蓄積する。

【0014】また、本発明の機械翻訳装置は、原言語単語と目的言語訳語候補の対訳関係の集合を保持する翻訳辞書と、該翻訳辞書を用いて翻訳対象の原言語単語列中の各原言語単語の目的言語訳語候補を検索する訳語候補生成手段と、該目的言語訳語候補と共起情報データベースのエントリとを照合して、原言語単語に対する該目的言語訳語候補対の共起強度を計算する共起強度検出手段と、前記共起強度を用いて、原言語単語に対する目的言語訳語列を選択する訳語決定手段を有する。

9

【0015】また、本発明の共起情報データベース構築装置は、目的言語の文の集合からなる目的言語コーパスを入力する目的言語入力手段と、該目的言語コーパス中の文において、定めた範囲内同時に出現する単語対とその共起頻度情報を共起情報データベースに蓄積する共起情報抽出手段を有する。

【0016】本発明の機械翻訳装置においては、訳語候補生成手段は、原言語単語と目的言語訳語候補の対訳関係の集合を保持する翻訳辞書を用いて、入力された翻訳対象の原言語単語列中の各原言語単語の目的言語訳語候補を検索する。共起強度検出手段は、目的言語の単語対とその共起頻度情報からなるエントリの集合を保持する共起情報データベースを用いて、原言語単語に対する目的言語訳語候補対の共起強度を計算する。訳語決定手段は、この共起強度の最も高くなる目的言語訳語候補の組み合わせを選び、原言語単語に対する目的言語訳語列を選択する。

【0017】これにより、原言語単語列に対して目的言語表現として実際に現れやすい最尤な訳語列を求めることが可能となり、本発明の目的である目的言語として自然な訳語列を選択することができるようになる。

【0018】また、目的言語入力手段は、目的言語の文の集合からなる目的言語コーパスを入力し、共起情報注出手段は、目的言語のコーパスを入力し、該目的言語コーパス中の文中で同時に出現する単語対とその共起頻度情報を前記共起情報データベースに蓄積する。これにより、目的言語共起情報データベースを目的言語コーパスから作成することが可能となり、共起情報データベースを用意するコストや手間が節約できる。

【0019】また、目的言語共起情報データベースは、目的言語の品詞タグ付き単語対とその共起頻度情報からエントリの集合を保持し、共起強度検出手段は、この品詞情報を含んだ共起情報データベースを利用して、品詞情報を利用して原言語単語に対する目的言語訳語候補対の共起強度を計算する。このとき、目的言語コーパスから共起情報データベースを得る場合は、各単語にその単語の品詞情報が付与された文の集合からなる目的言語コーパスを入力し、目的言語コーパス中の文において、定めた範囲内に同時に出現する品詞タグ付き単語対とその共起頻度情報を前記共起情報データベースに蓄積する。次に、共起強度検出手段において、品詞情報を含んだ共起情報データベースを利用して、品詞情報を利用して原言語単語に対する目的言語訳語候補対の共起強度を計算する。これにより、収集する目的言語単語の品詞を選択できたり、目的言語訳語候補中の目的言語単語とで、品詞による違いを正確に照合することが可能となる。

【0020】また、共起情報抽出手段は、品詞タグ付き単語対を収集する際に、あるキーとなる単語（キー単語）に対して共起する単語（共起単語）を品詞別に集計し、各共起単語について品詞別に何番目に近いかを示す

10

品詞別共起順位を抽出し、該キー単語と該共起単語の対、および該品詞別共起順位別に共起頻度情報を共起情報データベースに蓄積する。共起強度検出手段は、この品詞別共起順位と共起頻度情報を加味して共起強度を計算することが可能になる。これにより、依存関係を記述した目的言語コーパスを得るときに混ざる構文解析失敗などのノイズを回避することが可能で、依存関係の含まれない目的言語コーパスでありながら、共起情報を収集する際、依存関係のない単語対をある程度除外でき、少ないデータ量で有効な共起情報を得ることが可能となる。

【0021】また、分野入力手段において分野情報を入力後、分野コーパス抽出手段で、目的言語コーパスから該分野情報に関連する文の集合を抽出し、共起情報抽出手段は、該分野コーパス入力手段で抽出された文中で目的言語の共起頻度情報を収集する。これにより、最終出力結果の目的言語訳語列を指定の分野向きの訳語にチューンすることが可能となる。

【0022】また、共起強度検出手段は、各目的言語訳語候補と共起情報データベースのエントリ中の目的言語単語対を照合する際に、翻訳辞書および意味カテゴリ辞書を参照して、目的言語単語どうしだけでなく該目的言語単語の目的言語単語関連語および類義語とも照合して共起強度を計算する。これにより、同じ意味をもつ単語を表記に関係なく獲得が可能で、本発明の目的である単語の共起現象の希薄さの回避が可能となる。

【0023】また、翻訳辞書は、言語変換ルールによる訳語選択のルール制約情報を含み、訳語候補生成手段は、該翻訳辞書を参照して、目的言語訳語候補にルール制約情報を付与する手段を含む場合、訳語決定手段は、目的言語訳語候補に付与された該ルール制約に矛盾しない目的言語訳語候補を選択することができる。これにより、本発明の目的である原言語側のルール制約によって不要な目的言語訳語候補を除くことが可能となり、適当な訳語を選択する可能性を高めることができる。

【0024】また、訳語候補生成手段は、機械翻訳システムが持つ翻訳辞書に加えて、別の翻訳辞書を参照して原言語単語列中の各原言語単語の目的言語訳語候補を検索する手段を含む。これにより、目的言語訳語候補を増やし適当な訳語を選択する可能性を高めることができる。

【0025】また、原言語の単語対とその共起頻度情報からなるエントリの集合を保持する原言語共起情報データベースを用いて、原言語単語に対する原言語における共起強度をもとに、対応する訳語候補共起強度に重み付けを与えることが可能である。これにより、原言語で共起しやすい単語の訳語候補の共起関係を重視して訳語を選択することが可能となり、訳語選択精度をより向上させることができる。

【0026】

11

【発明の実施の形態】次に、本発明の実施形態について図面を参照して説明する。以下に示す実施形態では、原言語を日本語、目的言語を英語とする。

【0027】図1は、本発明の一実施形態の機械翻訳装置の構成を示す基本ブロック図、図2はその処理を示すフローチャートである。本機械翻訳装置は共起情報データベース構築部100と共起利用訳語選択部200と目的言語共起情報データベース300と辞書400と原言語共起情報データベース500より構成される。

【0028】共起情報データベース構築部100は分野入力部110と分野コーパス抽出部120と目的言語入力部130と共起情報抽出部140から構成される。分野入力部110で分野情報を入力し（ステップ610）、分野コーパス抽出部120で、目的言語の文の集合からなる目的言語コーパスから該分野情報に関連する文の集合を抽出する（ステップ620）。目的言語入力部130は目的言語コーパスを入力し（ステップ630）、共起情報抽出部140は目的言語コーパスの文において、定めた範囲に同時に出現する単語対とその共起頻度情報を目的言語共起情報データベース300に蓄積する（ステップ640）。¹⁰

【0029】辞書400は翻訳辞書410と意味カテゴリ辞書420から構成される。翻訳辞書410は原言語単語と目的言語訳語候補の対訳関係の集合を保持している。意味カテゴリ辞書420は目的語の単語とその単語の意味を代表する意味カテゴリとの対応関係の集合から構成される。

【0030】共起利用訳語選択部200は原言語単語入力部210と訳語候補生成部220と共起強度検出部230と訳語決定部240から構成される。訳語候補生成部220は、翻訳対象の原言語単語列を入力し（ステップ650）、翻訳辞書410を用いてその原言語単語列中の各原言語単語の目的言語訳語候補を検索する（ステップ660）。共起強度検出部230は、翻訳辞書410または意味カテゴリ辞書420を用いて各目的言語訳*³⁰

12

*語の関連語、類義語を検索し、共起情報データベース300を用いて原言語単語に対する訳語候補対の共起強度を計算する（ステップ670）。訳語決定部240は、共起強度を用いて、原言語単語に対する目的言語訳語列を選択する（ステップ680）。

【0031】次に、本実施形態における機械翻訳語選択の手順について説明する。ここでは、機械翻訳システムの入力文の一部が“有力市場調査機関”としたとき、これを構文解析した結果、図3に示すような依存関係になる。この依存関係を持った原言語単語列を入力例として、以下説明する。

【0032】まず、目的言語入力部130は目的言語コーパスを入力する。共起情報抽出部140は、一行コーパスからテキストを入力すると、定めた範囲内（例えば、読み込んだ一行内）で、同時に共起する単語対を全て走査し、共起情報データベース300内の該エントリの頻度を各々1増加させる。該当するエントリが存在しなければ、新たに頻度1のエントリとして追加する。この操作を繰り返す。

【0033】図4は形態素タグ付きの目的言語コーパスの例である。各単語は、“表記語/品詞タグ”で構成されている。例えば、“market/NN”は単数名詞の“market”を表す。最初の語“trading/NN”と共起する単語としては、（“trading/NN”、“stock-index/NN”）、（“trading/NN”、“futures/NNS”）、（“trading/NN”、“first/JJ”）、などが抽出できる。タグ付きコーパスから得られる形態素情報を用いることで、形態素タグの付いていない表記語の共起情報に比べて、後述の共起強度検出部230において、目的言語訳語候補中の目的言語単語と、より正確な照合が可能となる。表1は、この操作によって構築された共起情報データベース300の内容例である。

【0034】

【表1】

単語1	単語2	頻度
:	:	:
:	:	:
market/NN	organization/NN	3
market/NN	organizations/NNS	2
markets/NN	organization/NN	3
markets/NN	organizations/NNS	1
:	:	:
:	:	:

【0035】また、目的言語コーパスを入力する際に、分析する目的言語コーパスの分野を選択して共起情報データベース300を構築することもできる。まず、分野⁵⁰

入力部110で分野情報を入力する。例えば、「機械翻訳」などのキーワードを入力する。次に、分野コーパス抽出部120において、目的言語コーパスからその分野

13

情報に関連する文の集合を抽出する。例えば、大量のテキストデータから与えられたキーワードを含むテキストを抜き出すことのできる検索ソフトウェアを用いることで、目的とする分野に関連する文の集合が得られる。共起情報抽出部140では、この文集合を入力として共起情報データベース300を構築する。この共起情報データベース300を用いると、共起利用訳語選択部200で出力される目的言語訳語列をその分野向きのものにチューンすることが可能となる。

【0036】ここまでの操作は、共起利用訳語選択部200で利用することとなる共起情報データベース300を構築する処理であり、機械翻訳システムを実行する前に行っておく前処理となる。

*

見出し語	順位	訳語
有力だ	1	powerful
	2	strong
	3	leading
	4	influential
市場調査	1	market research
機関	1	organization
	2	means
	3	engine

【0040】このとき、機械翻訳システムが持つ翻訳辞書に加えて、別の翻訳辞書を参照することで、目的言語訳語候補を増やすことができる。図3の各原言語単語について、別の翻訳辞書を用いて対訳（表中、順位a、b※30

*【0037】次に、原言語単語入力部210は、機械翻訳システムの翻訳処理中から取り出した、翻訳対象の原言語単語列を入力する。依存関係を持った原言語単語列を図3に示す。

【0038】訳語候補生成部220は、原言語単語列の各原言語単語について、翻訳辞書400を検索し、目的言語の1語以上からなる目的言語訳語候補を得る。得られた目的言語訳語候補集合を対訳リストに保存する。図3の各原言語単語の対訳としては、例えば、表2に示す目的言語訳語候補が得られる。

【0039】

【表2】

※の単語）が追加された結果を、表3に示す。

【0041】

【表3】

見出し語	順位	訳語
有力だ	1	powerful
	2	strong
	3	leading
	4	influential
市場調査	1	market research
機関	1	organization
	2	means
	3	engine
	a	organ
	b	media

【0042】共起強度検出部230は、上記の処理で得られた訳語候補リストを用いて共起強度を抽出する。図5は、共起強度検出部230の処理を示すフローチャー

トである。以下では、原言語単語列のうち依存関係のある単語対に対する目的言語訳語候補対を扱う。

（ステップ701）まず、1つの目的言語訳語候補対に

15

含まれる訳語対を選ぶ。

(ステップ702) 次に、選んだ訳語対の各訳語に対して、変化形、派生語、同義語である目的言語単語関連語を考慮できる翻訳辞書410を参照して、この目的言語単語関連語を検索する。該関連語の組み合わせを展開 *

16

*し、その組み合わせのうちの関連語対を1つ選ぶ、表2の目的言語訳語候補の各訳語の関連語の集合を表4、表5に示す。

【0043】

【表4】

訳語	変化形
:	:
:	:
powerful	powerful/JJ
market	market/NN, markets/NNS
organization	organization/NN, organization/NNS, ...
:	:
:	:

【0044】

※ ※【表5】

訳語	変化形
:	:
:	:
powerful	powerful/JJ, powerfully/RB, powerfulness/NN, ...
market	market/NN, markets/NNS, market/VB, ...
organization	organization/NN, organization/NNS, organize/VB, organizational/RB, ...
:	:
:	:

【0045】また、関連語の組み合わせの検証が終わったら、訳語に対して同一の意味カテゴリを持つ類義語に対する組み合わせも考慮できる。すなわち、目的言語の単語とその単語の意味を表す意味素との対応関係の集合から構成される意味カテゴリ辞書600を用いて類義語を★

★検索し、該類義語の組み合わせも展開し、その組み合わせのうちの類義語対を1つ選ぶ。表2の各訳語の同一の意味カテゴリを持つ類義語の例を表6に示す。

【0046】

【表6】

意味カテゴリ	構成単語
:	:
:	:
(market)	market/NN, mart/NN, marketplace/NN, bazaar/NN, supermarket/NN, ...
(research)	research/NN, investigation/NN, search/NN, ...
:	:
:	:

【0047】以上2つの目的言語訳語候補の関連語を考慮することにより、目的言語コーパスを利用することによる単語共起現象の希薄さの問題を解決することが可能となる。

(ステップ703) 上記の共起情報データベース300から、選んだ関連語対のエントリと照合し、それが存在すれば、その共起頻度を取得する。

(ステップ704) 得られた共起頻度を、現在選んでい 50

る訳語対の頻度に加算する。ここで、頻度に対して関連語、類義語に応じてある係数をかけることにより、重み付けを行うこともできる。

(ステップ705) 訳語対に含まれる関連語、類義語について全て調べていれば、処理はステップ706に移行する。調べきれていなければ、処理はステップ701に移行する。表7に、訳語対の頻度を並べた訳語対頻度リストの例を示す。

【0048】

* * 【表7】

単語1	単語2	頻度
⋮	⋮	⋮
market	organization	9
research	organization	14
market	engine	5
research	engine	5
⋮	⋮	⋮

(ステップ706) 目的言語訳語候補対に含まれる訳語対について全て調べていれば処理はステップ707に移行する。調べきれていなければ処理はステップ701に移行する。

(ステップ707) 目的言語訳語候補対を1つ選ぶ。

(ステップ708) 上記で調べた、目的言語訳語候補対に含まれる全ての訳語対の頻度情報を用いて目的言語訳語候補対に対する共起強度を算出し、訳語候補共起強度リストに登録する。

【0049】共起単語対の頻度情報から、目的言語訳語※

※候補どうしの共起強度を算出する近似方法の一例を以下に示す。目的言語訳語候補対に含まれる各訳語対についての共起確率を計算する。訳語候補共起強度は、目的言語訳語候補に含まれる訳語対の組み合わせを考え、訳語対全ての共起確率の平均値とする。共起確率の代わりに、相互情報量などの他の統計値を用いてもよい。表8に、上記の処理によって得られた訳語候補共起強度リストを示す。

【0050】

【表8】

フレーズ1	フレーズ2	共起強度
⋮	⋮	⋮
market research	organization	0.3108
market research	engine	0.1351
market research	means	0.0405
leading	organization	0.0625
powerful	organization	0.125
⋮	⋮	⋮

(ステップ709) 目的言語訳語候補値について全て共起強度を算出していれば、共起強度検出部230の処理を終了する。調べきれていなければ、処理はステップ707に移行する。

【0051】次に、訳語決定部240は、上記訳語候補共起強度リストを用いて最終的な訳語列を選択する。図6は、訳語決定部240の処理を示すフローチャートである。(ステップ711) まず、訳語候補共起強度リストに訳語候補共起強度のエントリが残っていれば、処理はステップ712に移行する。リストが空であれば、訳語決定部240の処理を終了し、原言語単語の目的言語訳語候補の対応を最終出力とする。

(ステップ712) 訳語候補共起強度リストの中で、最も高い共起強度の値を持つエントリを検索し、取得す

る。

(ステップ713) 選択されたエントリの目的言語訳語候補を、対応する原言語単語の目的言語訳語候補として決定し、原言語単語と目的言語訳語候補の対応を保存する。

(ステップ714) 選択されたエントリを訳語候補共起強度リストから削除する。また、目的言語訳語候補が決定した原言語単語の目的言語訳語候補のうちで、ステップ713で選択された目的言語訳語候補以外の目的言語訳語候補を含むエントリを全て訳語候補共起強度リストから削除する。その後処理はステップ711に移行する。

【0052】ここでは、上の手順のもと目的言語訳語候補を選択していくと、まず、(“market research”,

19

“organization”の組が選択され、次に、“powerful”、“organization”が選択され、訳語候補共起強度リストが空となり、訳語決定部240の処理が終了する。最終的な原言語単語と目的言語訳語候補の対応は図8ようになる。

【0053】実際の機械翻訳システムでは、本発明の最終出力を用いて、各単語の語形変化、語順を整えて、目的言語の翻訳文が生成される。

【0054】また、共起情報抽出部230は、品詞タグ付き単語対を収集する際に、次のような工夫である程度*10

20

*依存関係のない単語対を除外することが可能となる。

【0055】図4を目的言語コーパスの例とすると、最初の語“trading/NN”をキーとなる単語（キー単語）としたとき、これと共起する単語（共起単語）を品詞別に集計し、それぞれの品詞別に何番目に近いかを示す品詞別共起順位を抽出する。表9にその集計結果の例を示す。

【0056】

【表9】

キー単語	品詞	共起順位			
		+1	+2	+3	...
trading/NN	N (名詞)	CBOE/NNP	stock-index/NN	futures/NNS	...
	V (動詞)	contract/VBP	be/VB	launched/VBN	...
	J (形容詞)	first/JJ	such/JJ	fast/JJ	...
:	:	:	:	:	:
:	:	:	:	:	:

【0057】次に、キー単語と共起単語との対、および 20 ※データベースを構築する。以上のようにして構築された共起品詞別共起順位別に共起頻度情報を共起情報データベース300に蓄積する。この操作を入力した目的言語コーパス中の全ての単語について行い、共起情報300データ※

情報データベース300の内容例を表10に示す。

【0058】

【表10】

キー単語	共起単語	品詞別共起順位			
		+1	+2	+3	...
:	:	:	:	:	:
:	:	:	:	:	:
market/NN	organization/NN	2	0	1	...
market/NN	organizations/NNS	2	0	0	...
markets/NN	organization/NN	2	1	0	...
markets/NN	organizations/NNS	1	0	0	...
:	:	:	:	:	:
:	:	:	:	:	:

【0059】さらに、共起強度検出部230は、この品詞別共起順位と共起頻度情報を加味して共起強度を計算する。例えば、共起順位のうち文末方向に1番目の共起頻度情報のみを用いて共起強度を計算することができる。 40

【0060】これにより、共起情報のデータベース300を構築するための資源が依存関係のない目的言語コーパスでありながら、依存関係のない単語対をある程度除外でき、少ないデータ量でも有効な共起関係を得ることができる。

【0061】また、機械翻訳システムへの入力文が、“有力市場調査機関が、・・・と予想している。”であるとしたとき、翻訳辞書410内のルール制約情報を検

索しマッチングさせると、図7の変換ルールが適用できると判断できる。図7が示す変換ルール中で、“（主体）”は原言語の意味カテゴリを示し、“主体”の意味を取り得る原言語単語が適用できることを示す。訳語候補生成部220は、このルール制約情報を目的言語訳語候補に付与する。つまり、原言語単語列中の単語“機関”の訳語はルールによって制約され、“organization”の訳語候補のみが“主体”の意味に取り得るとわかり、“organization”に制約“優先”の情報が付与される。ルール制約情報を付与した結果を表11に示す

【0062】

【表11】

21

22

見出し語	順位	訳語	制約
有力だ	1	powerful	—
	2	strong	—
	3	leading	—
	4	influential	—
市場調査	1	market research	—
機関	1	organization	優先
	2	means	制約
	3	engine	制約

【0063】次に、訳語決定部240において、ルールの制約条件が付いた目的言語訳語候補を優先的に採用する処理を行う。結果として、“機関”の訳語として“organization”が優先的に採用される。これにより、不要な目的言語訳語候補を早い段階で取り除くことが可能となる。

【0064】また、原言語における共起頻度情報からな*20

*るエントリの集合を保持するが原言語共起情報データベース500から得られる原言語共起情報を用いた優先訳語選択方法の一例を以下に説明する。原言語共起情報データベース500の例を表12に示す。

【0065】

【表12】

単語1	単語2	頻度
：	：	：
：	：	：
有力だ	機関	3
市場調査	機関	2
機関	予想する	3
：	：	：
：	：	：

【0066】原言語単語対の原言語における共起強度を、原言語単語列中の単語組み合わせの共起強度の和に対する該原言語単語対の共起頻度の割合とする。共起強度検出部230において、各訳語候補共起強度を、共起強度検出部230で計算した訳語候補共起強度と対応する原言語単語対の原言語における共起強度の積とする。*

※これを再び訳語候補共起強度リストに保存し、これを用いて訳語決定部240で最終的な訳語を選択する。表8が示す訳語候補共起強度リストを、以上の手順によって変更した結果を表13に示す。

【0067】

【表13】

フレーズ1	フレーズ2	共起強度
：	：	：
：	：	：
market research	organization	0.3108×0.4
market research	engine	0.1351×0.4
market research	means	0.0405×0.4
leading	organization	0.0625×0.6
powerful	organization	0.125×0.6
：	：	：
：	：	：

23

図9は本発明の他の実施形態の機械翻訳装置のブロック図である。

【0068】本実施形態の機械翻訳装置は入力装置801と記憶装置802～805と出力装置806と記録媒体807、808とデータ処理装置809で構成される。

【0069】入力装置801は目的言語コーパスおよび原言語単語列を入力するための、スキャナ、キーボードなどの入力装置である。記憶装置802、803、804はそれぞれ図1中の目的言語共起情報データベース300、言語共起情報データベース500、辞書400に相当する、記憶装置805はハードディスクである。出力装置806は目的言語訳語列が出力される、ディスプレイ、プリンタなどである。記録媒体807、808はFD（フロッピー・ディスク）、CD-ROM、MO（光磁気ディスク）などの記録媒体で、それぞれ図1中の共起データベース構築部100の各部の処理からなる共起データベース構築プログラム、共起利用訳語選択部200の各処理からなる共起利用訳語選択プログラムが記録されている。データ処理装置809は記録媒体807、808からそれぞれ共起データベース構築プログラム、共起利用訳語選択プログラムを記憶装置805に読み込んで、これらを実行するCPUである。

【0070】

【発明の効果】以上説明したように、本発明によれば、目的言語のコーパスを分析して得られる単語対とその共起頻度情報を共起情報データベースに蓄積し、この共起情報データベースを用いて、辞書によって検索された原言語単語に対する目的言語訳語候補対の共起強度を計算後、原言語単語に対する適切な目的言語訳語列を選択することにより、原言語単語列に対する最尤な訳語列を求めることが可能となる。分析する目的言語コーパスの分野を指定することにより、出力する目的言語訳語候補をその分野向きにすることが可能となる。

【0071】また、目的言語訳語候補に対する変化形、派生語、同義語、および同一意味カテゴリの類義語の考慮によって、共起現象の希薄さの問題点を解決することができる。

【0072】さらに、複数の翻訳辞書を利用することによる目的言語訳語候補の増加、原言語側のルール制約条件適用による不要目的言語訳語候補の除去、および原言語の単語共起情報の利用によって適切な訳語を選択する精度を高めることができる。

【0073】また、共起情報を抽出する際に、キーとなる単語と共起する単語、および品詞別の共起順位を抽出し、共起頻度情報とともに共起情報データベースに登録することにより、この品詞別共起順位と共起頻度情報を加味して共起強度を計算することが可能になり、これにより、共起情報を収集する際、構文解析失敗ノイズの回

24

避や、依存関係の含まれない目的言語コーパスでありながら、依存関係のない単語対をある程度除外することが可能となる。

【0074】以上のようにして、解析済みコーパスを用いることなく、共起情報データベースの規模を抑制しながら原言語単語列に対する最尤な訳語列を求めることが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施形態の機械翻訳装置の構成図である。

【図2】図1の機械翻訳装置の全体の処理を示すフローチャートである。

【図3】依存関係を持った原言語単語列の例を示す図である。

【図4】目的言語コーパスの例を示す図である。

【図5】共起強度検出部230の処理を示すフローチャートである。

【図6】訳語決定部240の処理を示すフローチャートである。

【図7】変換ルールの例を示す図である。

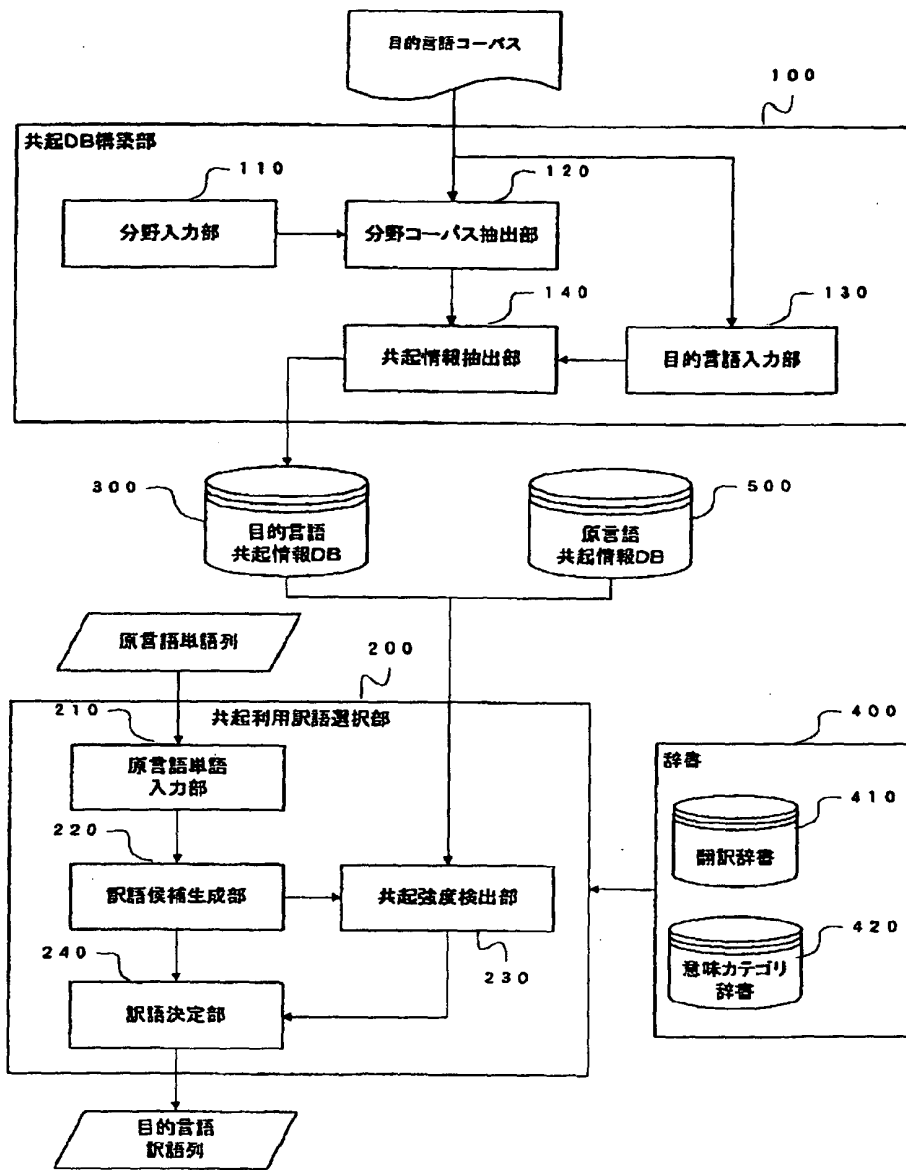
【図8】原言語単語と目的言語訳語候補の対応を示す図である。

【図9】本発明の他の実施形態の機械翻訳装置の構成図である。

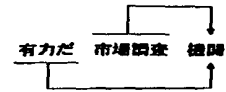
【符号の説明】

100	共起データベース構築部
110	分野入力部
120	分野コーパス抽出部
130	目的言語入力部
140	共起情報抽出部
200	共起利用訳語選択部
210	原言語単語入力部
220	訳語候補生成部
230	共起強度検出部
240	訳語決定部
300	目的言語共起情報データベース（共起情報データベース）
400	辞書
410	翻訳辞書
420	意味カテゴリ辞書
500	原言語共起情報データベース
610～680, 701～709, 711～714	ステップ
801	入力装置
802～805	記憶装置
806	出力装置
807, 808	記録媒体
809	データ処理装置

【図1】



【図3】



【図4】

Trading/NN in/IN CBOE/NNP 250/CD stock-index/NN futures/NNN ./, the/
OT first/JJ such/JJ futures/NNN contract/VBP to/TO be/VB launched/VBN
since/IN last/JJ year/NN 's/POS stock/NN market/NN crash/NN ./.

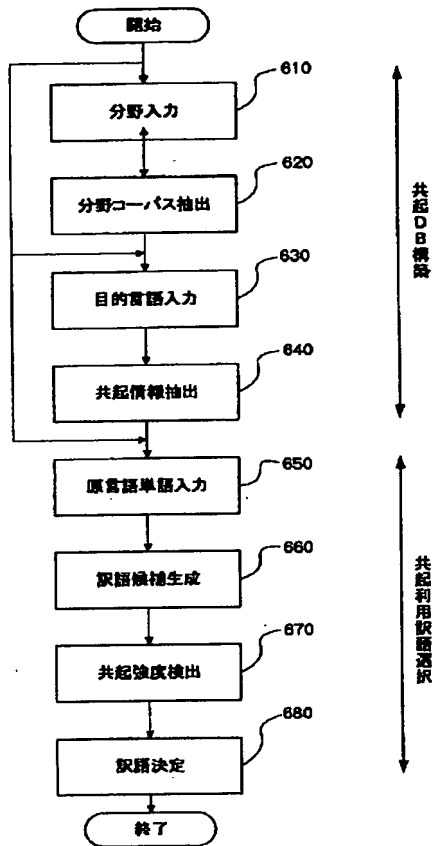
【図7】

((主体)が(事:抽象物)を予想する)

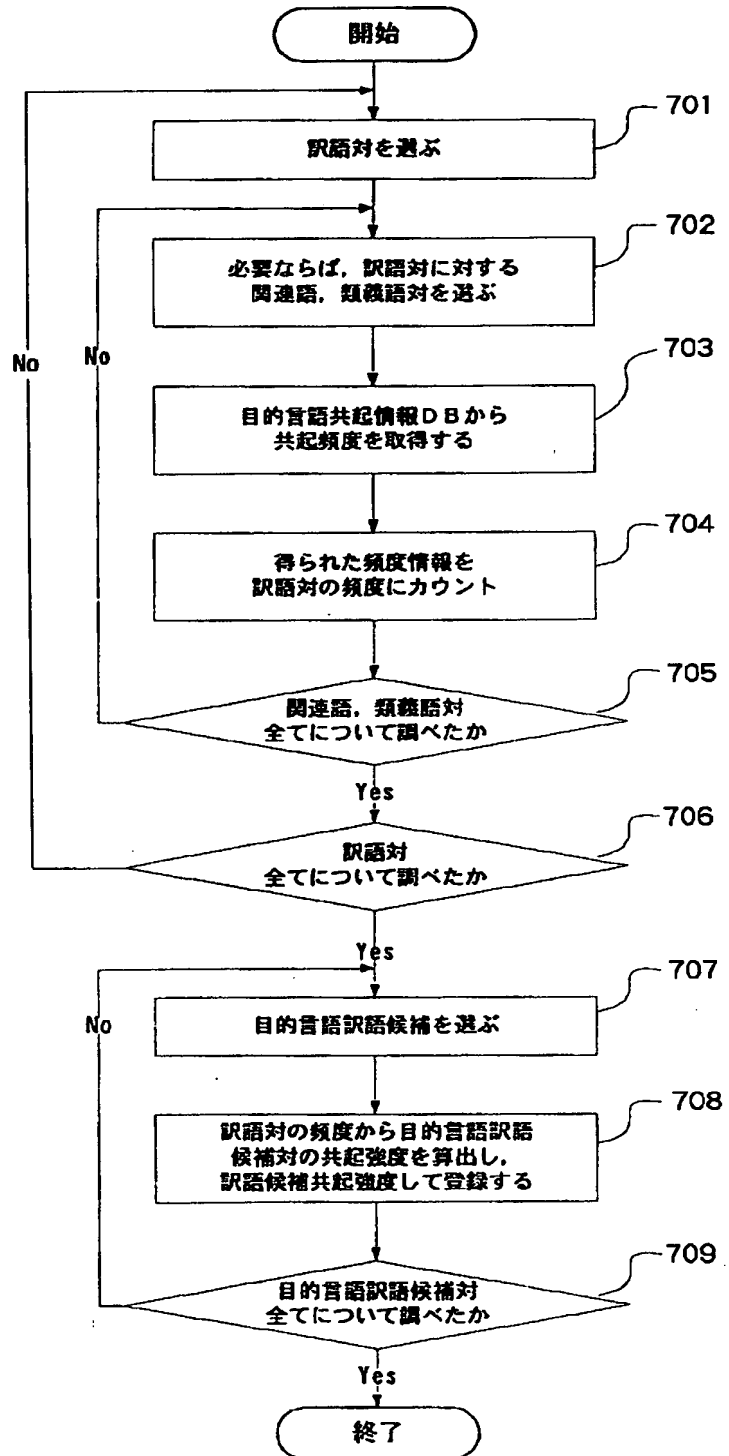
【図8】



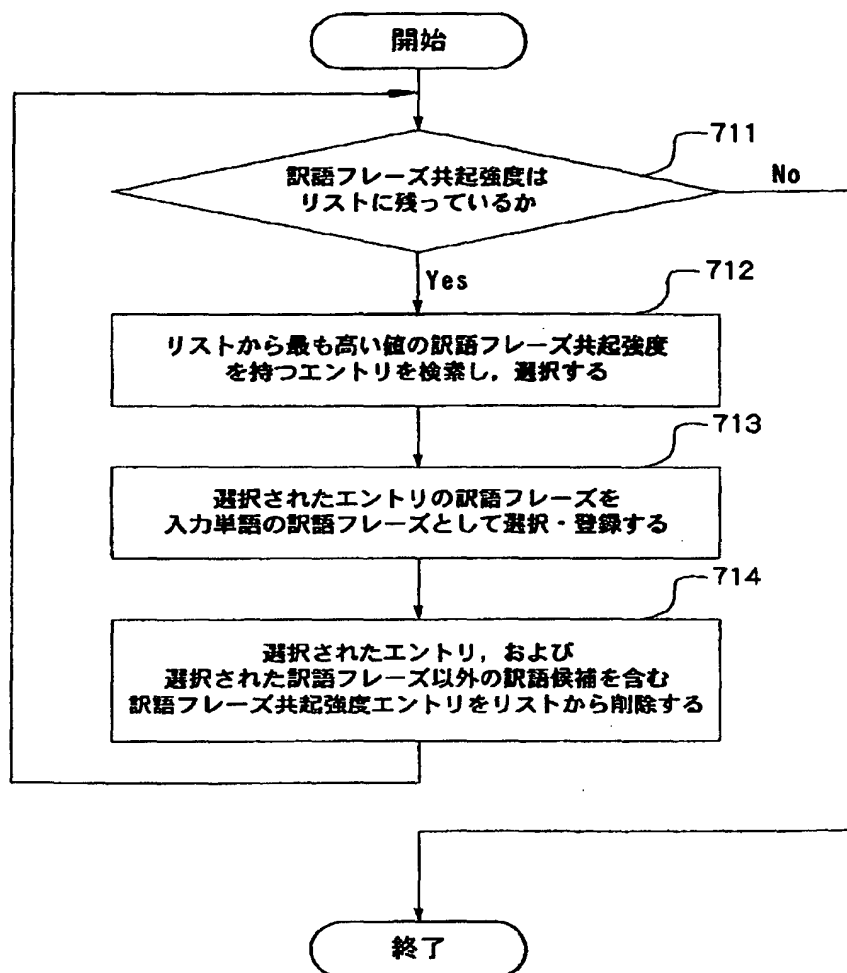
【図2】



【図5】



【図6】



【図9】

